Data in brief 28 (2020) 105069

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data regarding a new, vector-enzymatic DNA fragment amplification-expression technology for the construction of artificial, concatemeric DNA, RNA and proteins, as well as biological effects of selected polypeptides obtained using this method

Piotr M. Skowron [a, b, *, 1], Natalia Krawczun [a, b], Joanna Żebrowska [a, b], Daria Krefft [a, b], Olga Żołnierkiewicz [a], Marta Bielawa [b], Joanna Jeżewska-Frąckowiak [a, b], Łukasz Janus [a, b], Małgorzata Witkowska [a], Małgorzata Palczewska [a], Adriana Schumacher [c, d, 2], Anna Wardowska [d, e], Milena Deptuła [e], Artur Czupryn [f], Piotr Mucha [g], Arkadiusz Piotrowski [h, i], Paweł Sachadyn [j], Sylwia Rodziewicz-Motowidło [k], Michał Pikuła [d, e], Agnieszka Zylicz-Stachula [a, b, 1]

[a] Department of Molecular Biotechnology, Faculty of Chemistry, University of Gdansk, Gdansk, 80-308, Poland
[b] BioVentures Institute Ltd., Poznan, 60-141, Poland
[c] Department of Embryology, Faculty of Medicine, Medical University of Gdansk, Gdansk, 80-211, Poland
[d] Department of Clinical Immunology and Transplantology, Faculty of Medicine, Medical University of Gdansk, 80-210, Poland
[e] Laboratory of Tissue Engineering and Regenerative Medicine, Department of Embryology, Faculty of Medicine, Medical University of Gdansk, Gdansk, 80-211, Poland
[f] Nencki Institute of Experimental Biology, Warsaw, 02-093, Poland
[g] Department of Molecular Biochemistry, Faculty of Chemistry, University of Gdansk, 80-308, Poland
[h] Department of Biology and Pharmaceutical Botany, Faculty of Pharmacy, Medical University of Gdansk, 80-416, Poland
[i] International Research Agenda - 3P Medicine Lab, Medical University of Gdansk, 80-416, Poland

DOI of original article: https://doi.org/10.1016/j.msec.2019.110426.

* Corresponding author. Department of Molecular Biotechnology, Faculty of Chemistry, University of Gdansk, Gdansk, 80-308, Poland.

E-mail address: piotr.skowron@ug.edu.pl (P.M. Skowron).

[1] These authors contributed equally to this work.

[2] Present Address: [Adriana Schumacher], Department of Pharmacology, Faculty of Medicine, Medical University of Gdansk, Poland.

https://doi.org/10.1016/j.dib.2019.105069

[j] Laboratory for Regenerative Biotechnology, Faculty of Chemistry, Gdansk University of Technology, Gdansk, 80-233, Poland
[k] Department of Biomedical Chemistry, Faculty of Chemistry, University of Gdansk, Gdansk, 80-308, Poland

A R T I C L E I N F O

Article history:
Received 19 November 2019
Received in revised form 16 December 2019
Accepted 19 December 2019
Available online 31 December 2019

Keywords:
Concatemeric polypeptides
DNA amplification
DNA multimers
Peptide-based biomaterials

A B S T R A C T

Applications of bioactive peptides and polypeptides are emerging in areas such as drug development and drug delivery systems. These compounds are bioactive, biocompatible and represent a wide range of chemical properties, enabling further adjustments of obtained biomaterials. However, delivering large quantities of peptide derivatives is still challenging. Several methods have been developed for the production of concatemers – multiple copies of the desired protein segments. We have presented an efficient method for the production of peptides of desired length, expressed from concatemeric Open Reading Frame. The method employs specific amplification-expression DNA vectors. The main methodological approaches are described by Skowron et al., 2020 [1]. As an illustration of the demonstrated method's utility, an epitope from the S protein of Hepatitis B virus (HBV) was amplified. Additionally, peptides, showing potentially pro-regenerative properties, derived from the angiopoietin-related growth factor (AGF) were designed and amplified. Here we present a dataset including: (*i*) detailed protocols for the purification of HBV and AGF – derived polyepitopic protein concatemers, (*ii*) sequences of the designed primers, vectors and recombinant constructs, (*iii*) data on cytotoxicity, immunogenicity and stability of AGF-derived polypeptides.

1. Data description

We reported a method for efficient amplification of DNA fragments as linked in head-to-tail polymers, leading to the production of multiple copies of the desired peptide/peptides [1]. The data presented in this article demonstrate sequences of DNA primers and recombinant constructs used/obtained in the course of amplification, for the HBV epitope fragment and AGF-derived fragments. Data concerning selected biochemical properties and biological activity of the AGF-derived concatemeric proteins are also presented.

Fig. 1 shows a series of synthetic, amplifying DNA modules. The modules contain convergent recognition sequences for *Sap*I - a Type IIS restriction endonuclease (Rease), separated by a short DNA segment. The segment can contain an ancillary restriction site for *Sma*I for an alternative route of insert DNA cloning. Variants differ in terms of their possibility to manipulate three reading frames (which may be highly useful when amplifying natural, non-synthetic DNA sequences), as well as the presence or absence of a His6 tag. The amplifying modules may be re-cloned into various classes of DNA vectors, containing alternative origins of replication, antibiotic resistance genes, transcriptional promoters or translation signals. Following the amplification module, there are three stop codons in three reading frames. Genetic maps and DNA sequences of the constructed DNA amplification-expression vectors: pAMP1_A, pMAP1_B, pAMP1_C, pAMP1_HisA, pAMP1_HisB, pAMP1_HisC, pET28AMP-HisA are provided in Supplementary material in the GenBank file format. The pAMP vectors were designed on the

Specifications Table

Subject	Biomaterials
Specific subject area	Applying molecular biology methods for peptide-based biomaterials development
Type of data	Text files Figures Tables DNA sequences amino acid (aa) sequences
How data were acquired	flow cytometry, capillary electrophoresis, densitometry analysis, colorimetric assay, instruments: BD FACSCanto II flow cytometer, P/ACE MDQ System (Beckman Coulter, USA), Uncoated fused silica capillary (Postnova Analytics GmbH, Germany) software: UnScan-It (https://www.silkscientific.com), SnapGene version 4.1 (htt://www.snapgene.com), STATISTICA (StatSoft, Krakow, Poland).
Data format	processed data, analyzed data
Parameters for data collection	Cytotoxicity tests: Levels of lactate dehydrogenase (LDH) in culture media were measured after 48 hours from treatment with investigated proteins. Cell chemotaxis assay: migratory response for human cell lines was measured after 24 hours of incubation with the investigated proteins. *In vitro* immunogenicity assay: Peripheral blood mononuclear cells (PBMCs) were incubated with the investigated proteins for 48 hours. Selected immune cells populations: T helper cells – Th (CD3, CD4), T cytotoxic lymphocytes - CTL (CD3, CD8) and natural killer cells – NK (CD16, CD56) were measured with a flow cytometer. Basophil activation test (BAT): blood samples were incubated with the investigated proteins and stained with antibodies. Samples were analyzed through flow cytometry. Capillary electrophoresis: the plasma fraction from a blood sample was prepared and incubated with the investigated protein. Samples collected at different time points were analyzed with CE, using an uncoated fused silica capillary.
Description of data collection	Cytotoxicity tests: Cultures of 46BR.1 N fibroblasts and HaCaT keratinocytes were exposed to AGF_30 and AGF_40 of different concentrations. After 48 hours levels of realeased LDH were measured through a colorimetric assay. Cultures treated with Triton-X100 were used as a positive control. Cell chemotaxis assay: 46BR.1 N fibroblasts and HaCaT keratinocytes were cultured in silicone culture inserts in the presence of AGF_30 and AGF_40. Migratory response was measured after 24 hours. *In vitro* immunogenicity assay: PBMCs were incubated with the AGF_30 and AGF_40 concatemeric proteins for 48 h. After incubation, cells were stained with monoclonal, fluorochrome-labelled antibodies (anti-CD3, CD4, CD8, CD16, CD56, CD25, CD69, CD71, HLA-DR) and analyzed using a flow cytometer. Flow cytometry was performed in order to measure the level of activation in immune cell subpopulations. BAT test: Blood samples were incubated with the AGF_30 protein and AGF_40 protein and stained with the following antibodies: CD63, CD203c and CCR3. Samples were prepared for flow cytometry and measurements were taken. Capillary electrophoresis: the plasma fraction from a blood sample was prepared and incubated with AGF_30. CE analysis was performed for fractions collected at different time points. Measurements were taken using an uncoated fused silica capillary on a P/ACE MDQ System (Beckman Coulter, USA). DNA and aa sequences were created with SnapGene software. The intensity of DNA bands was measured with UnScan-It software. Statistical tests were performed using STATISTICA software.
Data source location	Gdansk, Poland
Data accessibility	With this article
Related research article	P.M. Skowron, N. Krawczun, J. Zebrowska, D. Krefft, O. Żołnierkiewicz, M. Bielawa, J. Jezewska-Frackowiak, L. Janus, M. Witkowska, M. Palczewska, A. Schumacher, A. Wardowska, M. Deptula, A. Czupryn, P. Mucha, A. Piotrowski, P. Sachadyn, S. Rodziewicz-Motowidlo, A. Zylicz-Stachula, A vector-enzymatic DNA fragment amplification-expression technology for construction of artificial, concatemeric DNA, RNA and proteins for novel biomaterials, biomedical and industrial applications, Mater. Sci. Eng. C 108 (2020), 110426. https://doi.org/10.1016/j.msec.2019.110426.

Value of the Data

- The presented dataset provides details of a novel method for efficient amplification of DNA fragments, in ordered, head-to-tail fasion, leading to the production of multiple copies of the desired peptide/peptides, linked in a single polypeptide
- The presented dataset provides detailed protocols for amplification, purification and immunodetection of AGF_30, AGF_40 and HBV protein S polymerized epitopes.
- This research shows effects of AGF_30 and AGF_40 on 46BR.1 N fibroblasts, HaCaT keratinocytes and peripheral blood mononuclear cells.
- These data are relevant for researchers interested in bioactive peptides production.
- Data show stability and anti-interference properties of peptides derived from the angiopoietin - related growth factor.

basis of p15A origin vector pACYC184 and pRZ4737. Thus, pACYC184 and pRZ4737 maps and DNA sequences are also included in Supplementary material.

Fig. 2 and Fig. 3 show DNA sequences of amplification cassettes with introduced alternations, enabling the purification and identification of the resulting polypeptides, such as secretion signals, tags or fusion proteins. The amplification cassette for fusion with ubiquitin is presented in Fig. 2. Both genetic map and sequence of the dedicated DNA vector pETAMP_SapI-Ubq are provided in Supplementary material in the GenBank file format. The amplification/secretion PhoA and MalE modules for secretion of the obtained recombinant concatemeric protein into the periplasmic space are presented in Fig. 3. Genetic maps and sequences of the dedicated DNA vectors: pETAMP_MalE and pET28AMP_PhoA are provided in Supplementary material in the GenBank file format.

Fig. 4 shows a diagram presenting the construction of the recombinant pAMP1-HisA_HBVepitope plasmid. The plasmid was constructed to assess the developed DNA fragment amplification technology [1]. A synthetic DNA fragment encoding a 7-aa epitope from the S protein of HBV was cloned into the *Sma*I restriction site of the amplification/expression vector pAMP1-HisA [Fig. 4]. The resulted recombinant plasmid pAMP1_HisA_HBVep_1 can be cleaved with *Sap*I to excise the modified DNA gene encoding the epitope. Then, the excised DNA fragment can be subjected to autoligation *in vitro*, yielding a series of DNA segments of increasing length, that are directional concatemers (polymers) of the epitope gene, as it was described by Skowron et al., 2020 [1]. The obtained concatemers can be recloned into the amplification-expression vector and subjected to another amplification cycle. The aa sequences of the poly-HBV epitopic proteins, obtained using this approach [1]: TKPTDGNGP_10, TKPTDGNGP_13, TKPTDGNGP_15, TKPTDGNGP_20, TKPTDGNGP_30 are included in Supplementary materials in FASTA format. The detailed maps and DNA sequences of the corresponding recombinant DNA constructs are also provided in the GenBank file format: pAMP1_HisA_HBVep_1, pAMP1_HisA_HBVep_10, pAMP1_HisA_HBVep_13, pAMP1_HisA_HBVep_15, pAMP1_HisA_HBVep_20, pAMP1_HisA_HBVep_30.

Fig. 5 shows an example of a synthetic, concatemeric gene, encoding the 25-mer HBV epitope. DNA sequence of the gene was optimized for expression in *Escherichia coli* (*E. coli*). Such a synthetic gene (the longest concatemeric DNA oligomer, obtainable by chemical synthesis), encoding the desired number of copies of the epitope, can be cloned into the *Sap*I-linearized DNA amplification/expression vector as initial DNA 'monomer'. The proposed alternative amplification approach may allow to obtain longer DNA concatemers in a shorter time [1]. The map and DNA sequence of the recombinant pAMP1_HisA_HBVep_25_opt, encoding the 25-mer HBV epitope is provided in Supplementary materials in the GenBank file format. The aa sequences of the obtained proteins poly-HBV epitopic proteins: TKPTDGNGP_opt_25, TKPTDGNGP_opt_50, TKPTDGNGP_opt_100, TKPTDGNGP_opt_150, TKPTDGNGP_opt_200, TKPTDGNGP_opt_450 [1] are included in Supplementary materials in FASTA format. The detailed maps and DNA sequences of the corresponding recombinant DNA constructs are also provided in Supplementary material in the GenBank file format: pAMP1_HisA_HBVep_25_opt, pAMP1_HisA_HBVep_50_opt, pAMP1_HisA_HBVep_100_opt, pAMP1_HisA_HBVep_150_opt, pAMP1_HisA_HBVep_200_opt, pAMP1_HisA_HBVep_450_opt.

The obtained recombinant DNA constructs, encoding various numbers of the HBV epitope, were used to transform *E. coli* BL21 Star (DE3) for concatemeric gene expression. Detailed protocols for purification of the poly-HBV epitope protein variants are presented in the Methods section of this article.

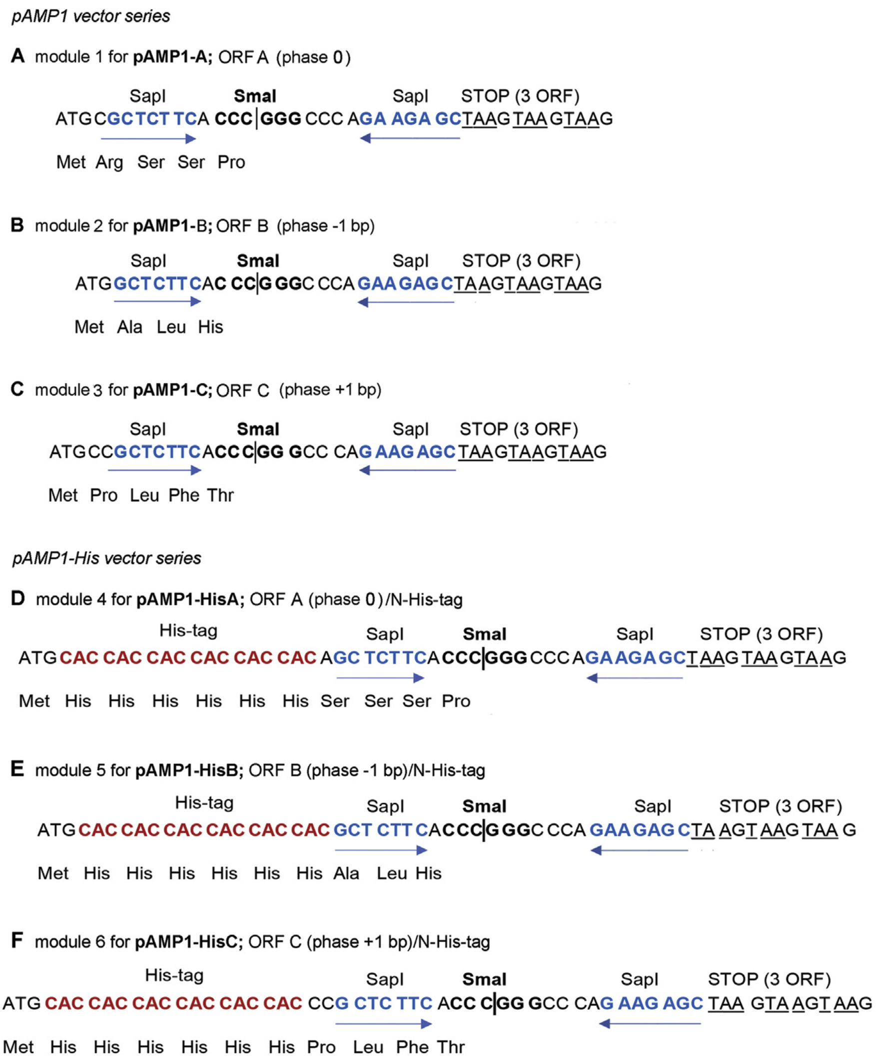


Fig. 1. A series of 6 modules designed for directional amplification of a DNA fragment maintaining ORF continuity. The modules contain convergent recognition sequences for *Sap*I - a Type IIS REase, separated by a short DNA segment. This segment can contain an ancillary restriction site for *Sma*I to provide an alternative route of insert DNA cloning. The variants differ in their possibility of manipulating three reading frames (which may be highly useful when amplifying natural, non-synthetic DNA sequences), as well as the presence or absence of a His6 tag. Variant D was used for the amplification of an epitope from the surface antigen S of HBV. The amplifying modules may be re-cloned into various classes of DNA vectors, containing alternative origins of replication, antibiotic resistance genes, transcriptional promoters or translation signals. Following the amplification module, there are three stop codons in three reading frames.

After the methods layout was established, DNA sequences encoding peptides of varying length, derived from the angiopoietin-related growth factor (AGF) were designed. The pET28AMP_MalE DNA vector [Supplementary material], enabling secretion of polyepitopic protein to the periplasmic space, was selected for cloning of genes encoding the AGF derivatives. DNA sequences of designed genes were optimized for expression in *E. coli*. An example of a designed gene coding for the 10-mer of the AGF-derived peptide is shown in Fig. 6. The aa sequences of the obtained AGF-derived proteins [1]: AGF_10, AGF_20, AGF_30 AGF_40 are provided in Supplementary materials in FASTA format. The detailed maps

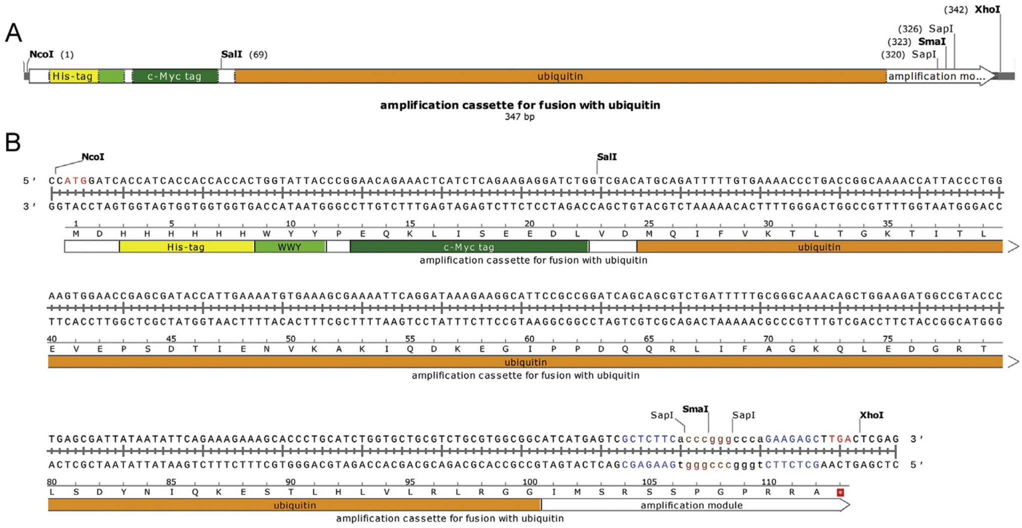


Fig. 2. Diagram showing the amplification cassette for N-terminal fusion of the poly-protein with ubiquitin. Panel A. A schematic representation of the designed module. Panel B. Nucleotide and aa sequences of the amplification/expression cassette along with the selected functional areas which contain DNA fragments encoding: His6 tag, c-Myc tag, WYY motif and ubiquitin. The WYY motif was introduced to allow a direct A280 concentration measurement of the concatemeric protein.

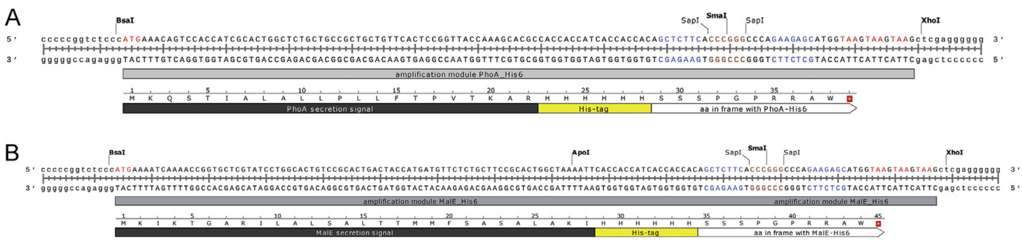


Fig. 3. Nucleotide and aa sequences of the amplification/secretion modules along with the selected functional areas which contain DNA fragments encoding: His6-tag and *E. coli* secretion signals. The designed modules enable secretion of the obtained recombinant concatemeric protein into the periplasmic space. Panel A. PhoA amplification module. Panel B. MalE amplification module.

and DNA sequences of the corresponding recombinant DNA constructs are provided in the GenBank file format: pET28AMP_MalE_AGF10, pET28AMP_MalE_AGF20, pET28AMP_MalE_AGF30, pET28AMP_MalE_AGF40.

Fig. 7 shows effectiveness of the exemplary autoligation reactions. The effectiveness of the reactions was determined by densitometric analysis of polyacrylamide gels, resulting from electrophoretic analysis of the DNA autoligation products, obtained in the first and second round of amplification of the model 7-aa HBV epitope [1, Figs. 3A and 4A]. The raw data corresponding to Fig. 7 are provided in Supplementary materials.

Figs. 8–11 show biological effects of the AGF-derived concatemeric proteins on human skin keratinocytes and fibroblasts. The cytotoxicity of the 30-mer and 40-mer of the AGF-derived peptide (AGF_30 and AGF_40, respectively) [Fig. 8], as well as, their influence on cell migration [Fig. 9], their immunogenicity [Fig. 10], and their allergenic potential [Fig. 11] were established as descried by

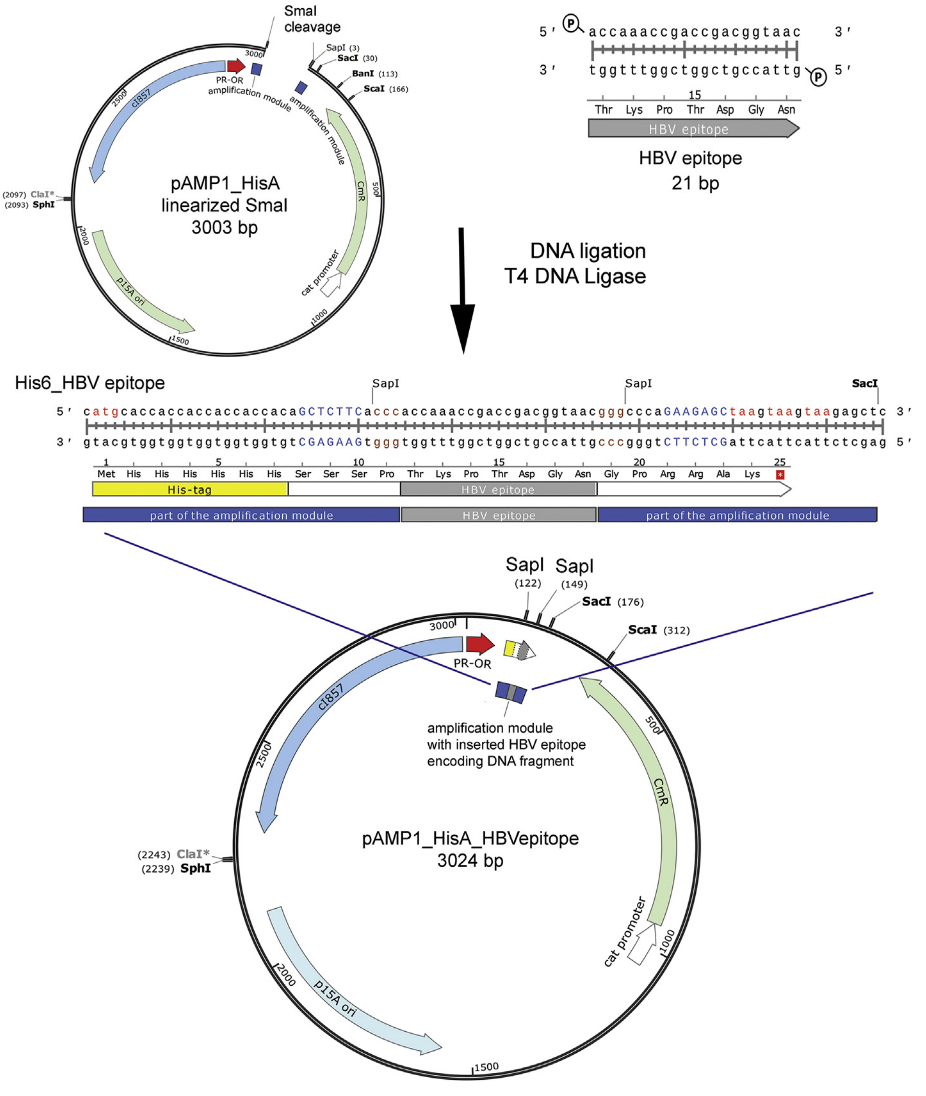


Fig. 4. Diagram showing the construction of the recombinant pAMP1-HisA_HBVepitope plasmid. A synthetic, phosphorylated, 21-bp DNA fragment (encoding the 7 aa epitope of HBV) was cloned into the *Sma*I restriction site of the amplification/expression vector pAMP1-HisA (Fig. 1D). The resulted recombinant plasmid can be cleaved with *Sap*I to excise the modified DNA gene encoding the epitope. Then, the excised DNA fragment can be subjected to *in vitro* autoligation, yielding a series of DNA segments of increasing length, which are directional concatemers (polymers) of the epitope gene. The obtained concatemers can be re-cloned into the amplification-expression vector and subjected to another amplification cycle. Alternatively, the obtained recombinant genes can be expressed in *E. coli* and the resulting poly-epitopic protein variants purified.

Skowron et al., 2020 [1]. The raw data corresponding to Figs. 8–11 are provided in Supplementary materials.

Fig. 12 shows stability of AGF_30 in human plasma. The protein stability was measured with capillary electrophoresis (CE) as descried by Skowron et al., 2020 [1]. Half-life (t 1/2 = 3.5 h) was

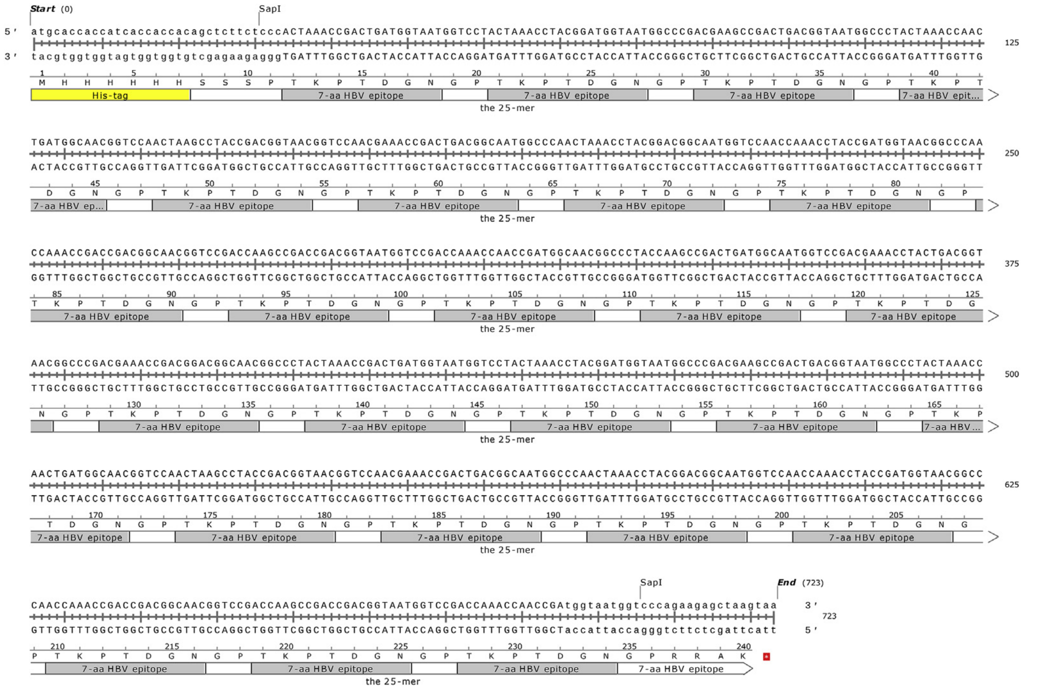


Fig. 5. DNA and aa sequences of the designed 25-mer of the 7-aa HBV epitope. The DNA gene sequence was optimized for expression in *E. coli*. The gene was chemically synthetized, cloned into the *Sap*I-cleaved pAMP1-HisA vector and subjected to the amplification process.

established, based on the change in area noted for the AGF_30 peak. The raw data are provided in Supplementary materials.

2. Experimental design, materials, and methods

2.1. Oligonucleotides

Primers for p15Aori-PR vector construction:	
PR-Sac-MCS	5′-CCCGAGCTCATATGGTCCATGGTCATGAACAACCTCCTTAGTACATGCAACCA-3′
PR-Cla	5′-GAAATCGATCGCGATCTTGCTCAATTGTTATCAGCTATG-3′
FpACYC-Sac	5′-GGCGCGAGCTCGACCCGGTCGAATTTGCTTTCGAATTTCTG-3′
RpACYC-ClaI	5′-GAATCTTCATCGATGGCATGCTGTCAAACATGAGAATTACAACT-3′
Primers for site-specific mutagenesis of pET21a(+) and pET28a(+)	
Fpet28adelsap	5′-CGTAAGGAGAAAATACCGCATCAGGCGCTCCTCCGCTTCCTCGCTCACTGACTCGCTG-3′
Rpet28adelSap	5′-CAGCGAGTCAGTGAGCGAGGAAGCGGAGGAGCGCCTGATGCGGTATTTTCTCCTTACG-3′
Fpet28adelsma	5′-GATATTCTTCTAATACCTGGAATGCTGTTTTaCCGGGGATCGCAGTGGTGAGTAACCATGCATC-3′
Rpet28adelsma	5′- GATGCATGGTTACTCACCACTGCGATCCCCGGtAAAACAGCATTCCAGGTATTAGAAGAATATC-3′
Primers for amplification of the insert from pAMP1-HisA recombinant constructs	
Fnest3	5′-CGTGCGTGTTGACTATTTTACCTC-3′
Rnest2	5′-CCCTTAAACGCCTGGTGCTACG-3′

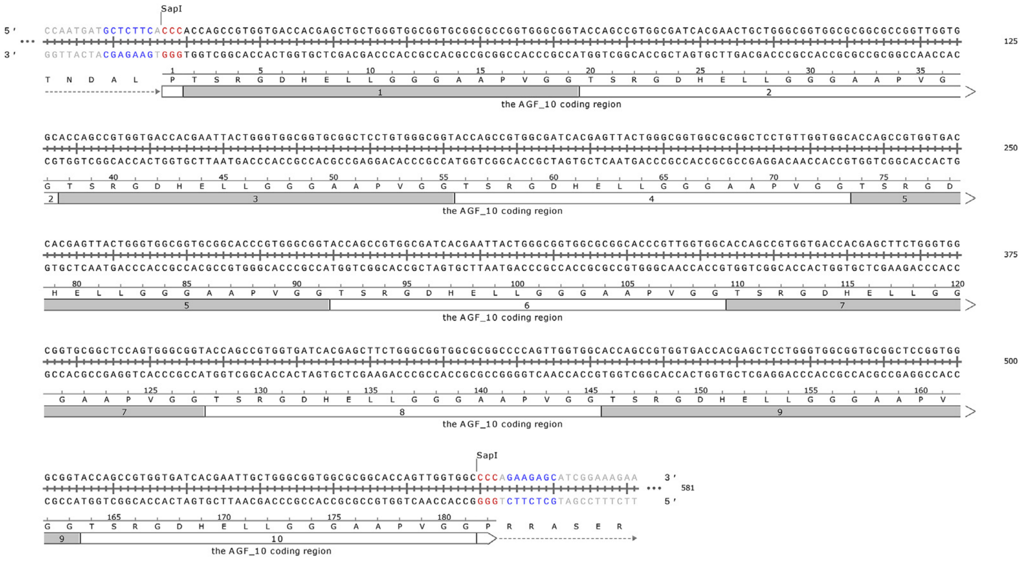


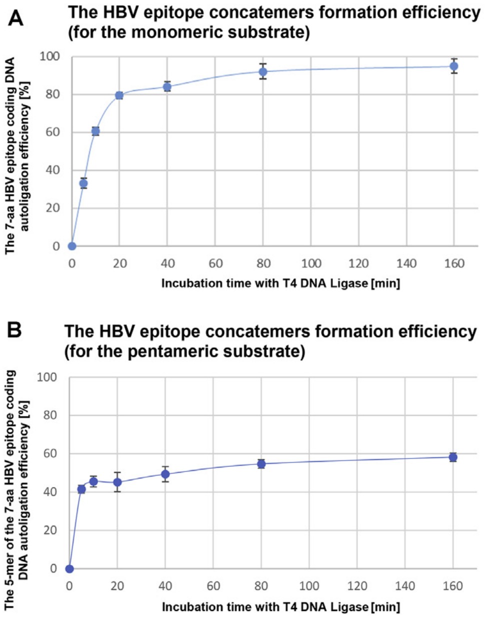
Fig. 6. Nucleotide and aa sequences of the designed 10-mer of the AGF-derived peptide. The DNA gene sequence was optimized for expression in *E. coli*. The gene was *de novo* synthetized and cloned into the *Eco*RV site of the pUCmini DNA vector. *Sap*I DNA recognition sequences are shown in blue. Nucleotides derived from pUCmini are marked in gray. *Sap*I cleavage results in the formation of 3-nt, single-stranded 5′ sticky ends. Apart from the amplification function, in the final concatemeric protein, these ends (shown in red) are responsible for the addition of a proline residue, the so-called "helical breaker", which separates epitope monomers and facilitates independent folding of the epitope into tertiary structures, thereby helping to maintain their natural spatial structure.

2.2. A detailed purification procedure of the HBV-derived polyepitopic protein variants

2.2.1. Resulting from the first and second amplification round

The purification scheme included the following stages:

1. *Sonication and heat treatment.* 1 L of bacterial culture was centrifuged (5000×g, 15 min). The bacterial pellet was resuspended in 30 ml of sonication buffer [50 mM Tris-HCl, pH 8.8 at 10 °C, 5 mM EDTA pH 8.0, 50 mM NaCl, 0.02% Triton X-100, 5 mM 2-mercaptoethanol (βMe), 1 × protease inhibitor solution (the Sigma FAST™ Protease Inhibitor), 0.5 mg/ml chicken egg lysozyme] and sonicated (8 × 30 s, 23 kHz, 3 μm). The lysate was centrifuged, and the supernatant was incubated at 70 °C for 30 min. Denatured thermolabile *E. coli* proteins were removed by centrifugation.
2. *Polyethyleneimine (PEI) treatment.* 10% PEI was gradually added to the supernatant to a concentration of 0.4%. After stirring for 30 min at 4 °C, the suspension was centrifuged.
3. *Ammonium sulphate (AmS) fractionation.* The resulting supernatant was subjected to two-step fractionation at 4 °C. At first, AmS was added to a final concentration of 30%. The suspension was incubated with stirring for 12 h and then centrifuged. During the second step, 50% saturation was applied, the suspension was incubated with stirring for 3 h and centrifuged. The resulting protein pellet was subjected to further purification steps.
4. *CM-Sepharose chromatography.* Cation exchange chromatography was conducted in buffer A [NaH_2PO_4/$NaHPO_4$ pH 6.0 at 10 °C] using a 1 ml CM-Sepharose ion exchange column and an automated Akta Pure chromatography system (GE Healthcare, Uppsala, Sweden), with a NaCl concentration gradient (0.7–1.1 M in buffer A).
5. *IMAC chromatography.* For IMAC purification, a 1 ml HiTrap® IMAC Fast Flow column (GE Healthcare) and the Akta Pure chromatography system were used. The protein sample was loaded onto the equilibrated Ni Sepharose resin (equilibration buffer: 20 mM Tris-HCl pH 8.0 at 10 °C, 0.5 M NaCl, 5



C

Lane (Skowron et al., 2020 [1], Figure 3A)	number of pixels in the DNA band	% of the monomer remaining	% of the monomer ligated into concatemers
K	1594283	100.00	0
1	1062022	66.61	33.39
2	632966	39.70	60.30
3	331200	20.77	79.23
4	246188	15.44	84.56
5	126914	7.96	92.04
6	103022	6.46	93.54
7 mixture of ligation products	721366	45.25	--

Lane (Skowron et al., 2020 [1], Figure 4A)	number of pixels in the DNA band	% of the monomer remaining	% of the monomer ligated into concatemers
K	2560528	100.00	0
1	1495349	58.40	41.60
2	1393854	54.44	45.56
3	1401267	54.73	45.27
4	1294865	50.57	49.43
5	1160248	45.31	54.69
6	1069141	41.76	58.24
7 mixture of the ligation products	1579313	61.68	--

Fig. 7. Formation efficiency of DNA concatemers coding for the 7 aa HBV epitopes. Panel A. Autoligation efficiency of the monomeric DNA substrate, encoding the 7 aa HBV epitope. Panel B. Autoligation efficiency of the pentameric DNA substrate, encoding five repeats of the 7 aa HBV epitope. The intensity of the DNA band corresponds to the monomer or pentamer measured using UnScan-It software; percentage of the non-ligated DNA substrate was calculated for each tested ligation time. Autoligation experiments were performed in triplicate. Panel C. Densitometric analysis of DNA bands corresponding to monomeric (see Fig. 3A, Skowron et al., 2020 [1]) and pentameric (see Fig. 4A, Skowron et al., 2020 [1]) DNA substrates.

mM imidazole).Then, the resin was washed with 5 column volumes (CV) of equilibration buffer. Additional washing was performed with 5 CV of washing buffer (20 mM Tris-HCl pH 8.0 at 10 °C, 0.5 M NaCl, 20 mM imidazole). Proteins were stepwise eluted with 6 CV of elution buffer (20 mM Tris-HCl pH 8.0 at 10 °C, 0.5 M NaCl, 250 mM imidazole).

6. *Resource S chromatography.* Cation exchange chromatography was conducted in buffer A [20 mM Tris-HCl pH 6.0] using a 6 ml Resource S ion exchange column and the Akta Pure chromatography system, with a NaCl concentration gradient (0.7–0.9 M in the buffer A).

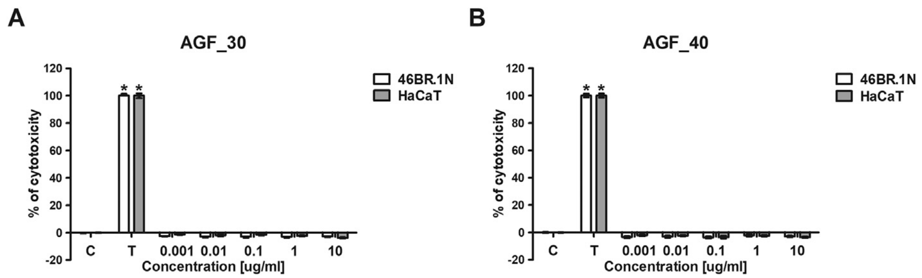


Fig. 8. Cytotoxicity of the protein variants AGF_30 (TSRGDHELLGGAAPVGG_30) and AGF_40 (TSRGDHELLGGAAPVGG_40) towards 46BR.1N fibroblasts and HaCaT keratinocytes. Results are presented as the mean ± SEM. All results come from four independent experiments performed in quadruple. *$p < 0.05$ relative to control (C), U-Whithey Test. T-additional positive control, cells treated with Triton X100 (1.0%) - maximum LDH release (maximum cytotoxicity).

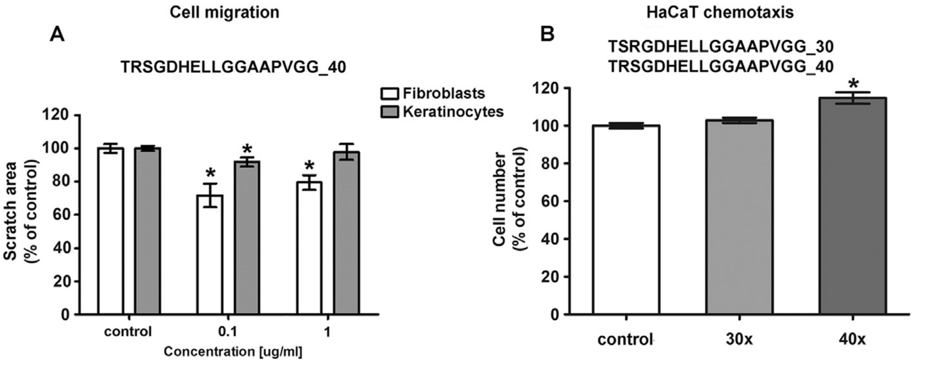


Fig. 9. Effects of AGF-derived poly-protein variants on cell migration. Panel A. The migratory response of human 46BR.1 N fibroblasts and HaCaT keratinocytes to the AGF_40 poly-protein after 24 h of incubation. Panel B. Effect of the AGF-derived poly-proteins on chemotactic migration of the human keratinocyte cell line (HaCaT) after 24 h of incubation. The final protein concentration was 1 μg/ml. All results were obtained from four independent experiments performed in quadruple and presented as mean ± SEM. *p < 0.05 relative to control, U-Whithey Test.

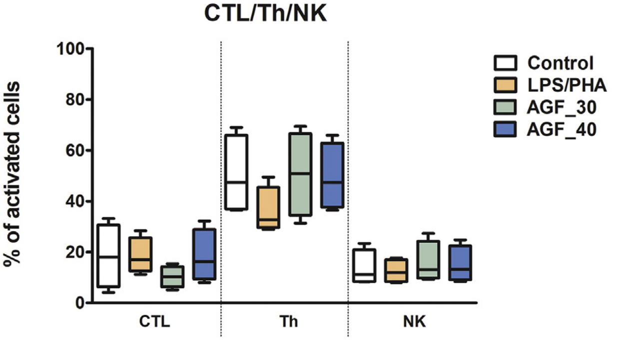


Fig. 10. Immunogenicity of the AGF-derived poly-proteins. The *in vitro* immunogenicity assay was performed on human PBMCs exposed to the AGF-derived polyprotein variants AGF_30 (TSRGDHELLGGAAPVGG_30) and AGF_40 (TSRGDHELLGGAAPVGG_40) for 48 h. The final protein concentration was 1.0 μg/ml. The analysis was performed with flow cytometry in order to evaluate the activation level of selected immune cell subpopulations. The chart presents the overall activation percentage of selected lymphocytes/leukocytes: CTL - T cytotoxic lymphocytes, Th − T helper cells, NK- natural killer cells. Results were obtained from four independent experiments performed and are presented as medians with min = max. p − value was evaluated with the Mann-Whitney *U* Test (in comparison to positive control). Untreated cells constituted a negative control, whereas cells stimulated with LPS (1 μg/ml, lipopolysaccharide) and PHA (2.5 μg/ml, phytohemagglutinin) were a positive control.

7. *Size exclusion* chromatography (SEC). SEC was performed using an Akta Pure chromatography system and HiLoad 16/600 Superdex 200 PG column, in buffer S [20 mM Tris-HCl pH 8.0 at 10 °C, 100 mM NaCl, 0.1 mM EDTA]. The purified protein preparations were concentrated, dialysed against PBS buffer and lyophilized.

2.2.2. *Resulting from the alternative amplification round*

The purification scheme included the following stages:

1. *Sonication and heat treatment.* Performed as described in a).
2. *PEI treatment.* Performed as described in a).
3. *AmS fractionation.* Performed as described in a), except that 60–70% AmS saturation was applied.

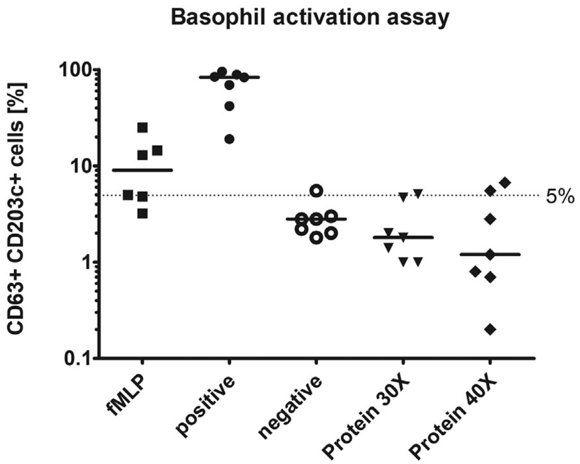


Fig. 11. ***In vitro*** **activation of basophils in the presence of activating antibodies (first positive control), fMLP (second positive control), the AGF_30 protein, the AGF_40 protein (1 μg/ml) and negative control.** Results are presented as a single data (% of activated basophils for each patient) and median.

4. *IMAC chromatography.* For IMAC purification, a 5 ml HiTrap® IMAC Fast Flow column (GE Healthcare) and the Akta Pure chromatography system were used. The protein sample was loaded onto the equilibrated Ni Sepharose resin (equilibration buffer: 20 mM Na/PO_4, 500 mM NaCl, pH 7.4 at 10 °C).Then, the resin was washed with 5 column volumes (CV) of equilibration buffer. Proteins were eluted with imidazole gradient (0−500 mM the equilibration buffer).
5. *Resource Q chromatography (used as a negative step).* Anion exchange chromatography was conducted in buffer Q [20 mM Tris-HCl, 0.1 mM EDTA, pH 9.0 at 10 °C] using a 6 ml Resource Q ion exchange column and the Akta Pure chromatography system. The protein sample was loaded onto the equilibrated resin. Then, the resin was washed with 10 CV of equilibration buffer. In the buffer conditions applied, the polyepitopic proteins did not bind to the resin and eluted from the column in the wash fraction. Purified protein preparations were concentrated and subjected to SEC.
6. *SEC* chromatography. SEC was performed using an Akta Pure chromatography system and HiLoad 16/600 Superdex 200 PG column, in buffer S [20 mM Tris-HCl pH 8.0 at 10 °C, 100 mM NaCl, 0.1 mM EDTA]. Purified protein preparations were concentrated, dialysed against PBS buffer and lyophilized.

2.3. *A detailed purification procedure of the AGF-derived poly-protein variants*

The purification scheme of the AGF-derived poly-protein variants from *E. coli* BL21(DE3)* cells included the following stages:

1. *Lysis.* In order to extract periplasmically secreted proteins, the bacterial cell pellet was resuspended in 100 ml of sucrose buffer [20% sucrose, 5 mM EDTA]. After 30 min incubation at room temperature (RT), the cells were centrifuged, and the resulting cell debris was resuspended in 60 ml of pre-cooled 10 mM $MgCl_2$. After 10 min incubation on ice, the lysate was centrifuged.
2. *AmS fractionation.* The resulting supernatant was subjected to two-step fractionation at 4 °C. At first, AmS was added to a final concentration of 30%. The suspension was incubated for 1 h with stirring, and then centrifuged. During the second step, 60% saturation was applied, the suspension was

incubated with stirring for 1 h and centrifuged. The resulting protein pellet was subjected to further purification steps.

3. *IMAC chromatography.* The pellet was resuspended in 110 ml of buffer A [50 mM NaH_2PO_4, 500 mM NaCl] and filtered. Chromatography was performed in buffer A using a 5 ml HiTrap® IMAC Fast Flow column (GE Healthcare) and the Akta Pure chromatography system (GE Healthcare, Uppsala, Sweden), with an imidazole concentration gradient (10–500 mM in the buffer A). Chromatographic fractions containing the AGF-derived protein variant were collected and dialyzed against 1x PBS buffer at 4 °C.
4. *Affinity based lipopolysaccharide (LPS) removal.* Endotoxin removal was performed in accordance with the manufacturer's instructions. 20 ml of the protein sample was transferred into the equilibrated Proteus NoEndo M spin column and centrifuged for 30 min at 100×g. The resulting, purified protein preparation was free of LPS.

2.4. LC–MS–MS/MS analysis

Selected homogenous protein variants were subjected to liquid chromatography coupled to tandem mass spectrometry (LC–MS–MS/MS) [2]. LC-MS-MS/MS analysis were performed in a Mass Spectrometry Laboratory (IBB PAS, Warsaw). Gel slices containing the protein variant were subjected to a standard 'in-gel digestion' procedure. Any protein disulphide bonds were reduced with 100 mM DTT (30 min at 56 °C), alkylated with iodoacetamide (45 min; in a darkroom; room temperature) and digested overnight with trypsin (sequencing Grade Modified Trypsin, Promega). The resulting peptides were eluted from the gel with 0.1% trifluoroacetic acid (TFA) and 2% acetonitrile (ACN) and measured by LC/MS. HPLC separation parameters: precolumn: RP-18 (nanoACQUITY Symmetry R ® C18,Waters), 0.1% TFA as a mobile phase. Nano-HPLC: RP-18 column (nanoACQUITY BEH C18, Waters), flow rate 250 nl/min, gradient: 0–35% B in 70 min, solvent A: 0.05% formic acid in water, solvent B: 0.05% formic acid in ACN. The column outlet was directly coupled to the ion source of the spectrometer working in the regime of data dependent MS to MS/MS switch (Orbitrap Velos mass spectrometer-Thermo Electron Corp.). Raw data were processed using Mascot Distiller followed by Mascot Search (Matrix Science, UK) against the predicted protein-derived reference peptide masses. Search parameters selected for LC–MS–MS/MS for the precursor and product ions mass tolerance were as follows: 20 ppm and 0.1 Da, respectively. Trypsin specificity allowed for one missed cleavage site. Fixed modification selected for the search: cysteine by carbamidomethylation. Variable modification selected for the search: oxidation of methionine. Peptides with a Mascot Score exceeding the 5% False Positive Rate threshold and with a Mascot Score exceeding 30 were positively identified.

2.5. Western blotting and immunodetection procedure

The purified protein variants were separated by SDS-PAGE in 15% Tris-Glycine gels and electroblotted onto a PVDF membrane, using a semi-dry Trans-Blot Turbo Transfer System and a ready-to-use preassembled Trans-Blot Turbo Mini PVDF Transfer Pack at 2.5 mA, 25 V for 10 min in Towbin buffer [25 mM Tris, 192 mM glycine, pH 8.3 at 25 °C], with 20% methanol and 0.02% SDS. The membrane was blocked for 2 h 4 °C in TBS-T buffer [50 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20, pH 7.6 at 25 °C] with 3% skimmed milk. The membrane was then incubated with conjugated anti-His-HRP antibodies diluted 1:2000 in TBS-T buffer with 3% skimmed milk for 1 h at 30 °C. The membrane was washed three times with TBS-T buffer and the specific protein was visualized by adding a DAB solution.

2.6. Statistical analysis

Where statistical analyses were perfomed, data are given as mean values ± standard error of the mean (SEM). The number of independent experiments is indicated in the Figure legends as n. Statistical

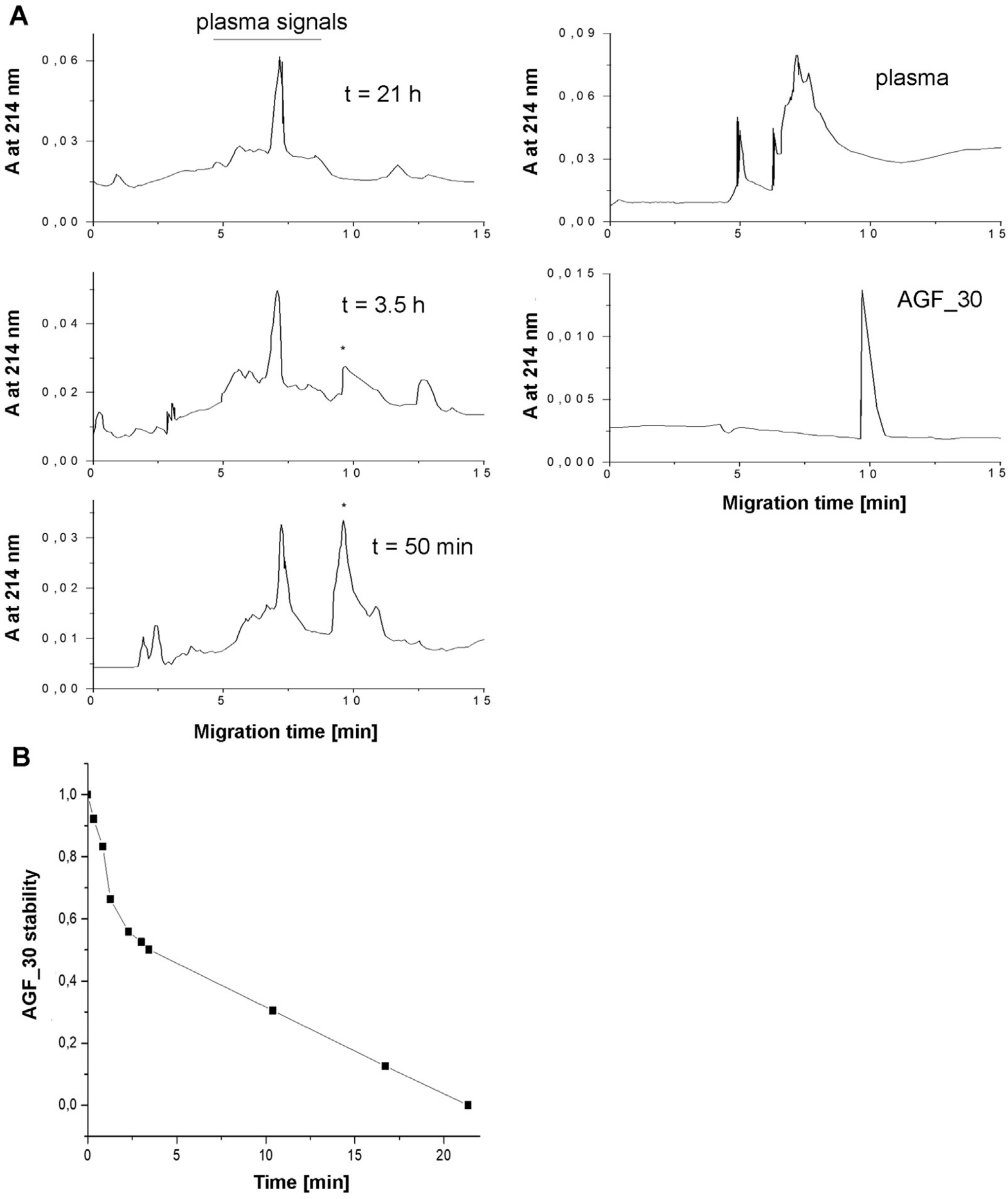


Fig. 12. The AGF_30 protein stability in human plasma. Panel A. Capillary electrophoresis (CE) analysis of AGF stability in human plasma. Panel B. Stability profile of AGF_30 based on the obtained CE data. Separation conditions: uncoated fused silica capillary of 40 cm (30 to detector) × 75 μm; background electrolyte (BGE): 25 mM phosphate (pH 7.0), 50 mM SDS; separation at 20 kV; temperature 25 °C; hydrodynamic injection at 0.5 psi for 8 s; normal polarization; detection at 214 nm. The AGF_30 peak is marked with an asterisk. Incubation time of AGF_30 with human plasma is shown.

comparisons were performed with STATISTICA (StatSoft, Krakow, Poland), the Mann-Whitney *U* Test was used and P values are indicated as $^{*}p < 0.05$.

2.7. Evaluation of the autoligation efficiancy

The autoligation of DNA fragments (encoding either the 7aa HBV epitope monomer or pentamer) was performed as described by Skowron et al., 2020 [1]. The autoligation products were separated using polyacrylamide or agarose gel electrophoresis. Polyacrylamide and agarose gels were prepared in 1 × Tris-Borate-EDTA (TBE) buffer. The DNA bands were visualized after staining with Sybr Green I using a 312-nm UV transilluminator and photographed with a SYBR Green gel stain photographic filter.

The autoligation reaction effectiveness was determined by densitometry analysis of the selected DNA bands. The intensities of DNA bands, corresponding to the 7aa HBV epitope monomer or pentamer were measured using UnScan-It software and the persentage of the non-ligated DNA substrate was calculated for each ligation time tested. Autoligation experiments were performed in triplicate.

Funding

This work was supported by the National Centre for Research and Development – Poland [grant numbers POIG.01.04.00-22-140/12 to BioVentures Institute Ltd. (Poznan, Poland), STRATEGMED1/235077/9/NCBR/2014 to University of Gdansk, Faculty of Chemistry, Molecular Biotechnology Department] and BioVentures Institute Ltd. Funding for open access charge: University of Gdansk. The technology is owned by BioVentures Institute Ltd. (Poland) and protected by a Polish patent no 228341 [3] (filed in 2014), PCT patent application filed in 2015 and the global patent applications [4] which are currently being processed. Relevant patent (concerning the AGF-derived protein variants) pending: P.427146 [5].

Acknowledgments

We thank Prof. Mirosława Cichorek (Department of Embryology, Medical University of Gdansk), Prof. Janusz Moryś (Chairperson of Anatomy, Medical University of Gdansk, Poland) and Prof. Piotr Trzonkowski (Department of Medical Immunology, Medical University of Gdansk) for granting access to equipment. We would also like to thank Maciej Zielinski, PhD and the personnel of the Laboratory of Immunology and Clinical Transplantology from The University Clinical Centre in Gdansk for carrying out BAT assays. We thank Agnieszka Ożóg for excellent project administration.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.105069.

References

[1] P.M. Skowron, N. Krawczun, J. Zebrowska, D. Krefft, O. Zołnierkiewicz, M. Bielawa, J. Jezewska-Frackowiak, L. Janus, M. Witkowska, M. Palczewska, A. Schumacher, A. Wardowska, M. Deptula, A. Czupryn, P. Mucha, A. Piotrowski, P. Sachadyn, S. Rodziewicz-Motowidlo, A. Zylicz-Stachula, A vector-enzymatic DNA fragment amplification-expression technology for construction of artificial, concatemeric DNA, RNA and proteins for novel biomaterials, biomedical and industrial applications, Mater. Sci. Eng. C 108 (2020) 110426, https://doi.org/10.1016/j.msec.2019.110426.

[2] D. Hess, T.C. Covey, R. Winz, R.W. Brownsey, R. Aebersold, Analytical and micropreparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis, Protein Sci. 2 (1993) 1342–1351.

[3] P. Skowron, A. Żylicz-Stachula, O. Żolnierkiewicz, M. Skowron, Ł. Janus, J. Jeżewska-Frąckowiak, A. Szymańska (2018) Patent PL 228341-B1.

[4] P. Skowron, A. Zylicz-Stachula, O. Zolnierkiewicz, M. Skowron, L. Janus, J. Jezewska-Frackowiak, D. Krefft, D. Nidzworski, K. Szemiako, N. Maciejewska, M. Nowak, A. Szymanska (2015) PCT Patent Application WO201 5162560 A1.
[5] P. Skowron, A. Żylicz-Stachula, J. Żebrowska, M. Palczewska-Groves, N. Maciejewska, A. Czupryn, Ł. Janus, P. Mucha, M. Deptuła, A. Piotrowski, M. Pikuła, S. Rodziewicz-Motowidło, J. Sawicka. (2018) Republic of Poland Patent Office, Patent Application P.427146.